Opinion paper for submission to Food Quality and Preference

Trained vs. consumer panels for analytical testing: Fueling a long lasting debate in the field

Gastón Ares[1] & Paula Varela*[2]

[1] Sensometrics & Consumer Science, Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República. By Pass de Rutas 8 y 101 s/n. C.P. 91000. Pando, Canelones, Uruguay

[2] Consumer and Sensory Sciences. Nofima AS, P.O. Box 210, 1431 Ås, Norway

* Corresponding author: Paula Varela [Email: paula.varela.tomasco@nofima.no]

Abstract

Sensory evaluation has traditionally been divided into two clearly defined areas: analytical tests, aimed at objectively evaluating the sensory characteristics of products, and hedonic tests, in which consumers evaluate their acceptance/preference. One of the central dogmas of the field has been matching these two types of tests to different types of assessors respectively: selected and trained assessors and regular consumers of the target products. Consumers have been for years regarded as not capable of performing analytical tasks. However, the development of various alternative methods for sensory characterization in the last couple of decades, has agitated the debate about the use of untrained assessors for analytical tasks in sensory science. Lately, the line between trained and consumer panels for analytical tests has blurred and is expected to continue to do so. The present opinion paper discusses some of the most relevant issues around the debate of whether consumer or trained assessor panels are appropriate for analytical testing in specific application and to provide recommendations for practitioners on this respect.

Keywords: *descriptive analysis; sensory characterization; sensory evaluation; trained assessors; consumers; panels*

1. Introduction

Sensory evaluation can be defined as a scientific discipline that evokes, measures, analyzes, and interprets responses to the characteristics of products as perceived by the senses (Stone & Sidel, 2004). This discipline has traditionally been divided into two clearly defined areas: analytical tests, aimed at objectively evaluating the sensory characteristics of products, and hedonic tests, in which consumers evaluate their acceptance/preference (O'Mahony, 1995). One of the central dogmas of the field has been matching these two types of tests to different types of assessors (Lawless & Heymann, 2010).

Analytic tests have traditionally been performed with trained assessors, who are selected based on their sensory acuity for basic characteristics (basic tastes, odours and textures) and their ability to discriminate among products (Stone & Sidel, 2010). After selection, assessors are familiarized with the testing procedures and are trained and retrained to recognize/describe/quantify the sensory characteristics of the target products in a reliable way (Lawless & Heymann, 2010).

On the other hand, hedonic tests are carried out with frequent consumers of the target products, which are asked to indicate their liking or preference based on an integrated evaluation (Lawless & Heymann, 2010). Consumers perceive products as a whole and usually give different relative importance to the sensory characteristics of products through a process of synthesis that determines their hedonic reaction (Jaeger, Wakeling, & MacFie, 2000).

The distinction between analytic and hedonic tests implies that test methods and assessors cannot be mismatched (Lawless & Heymann, 2010). Wide consensus exists regarding the idea that trained assessors cannot perform hedonic tests, as they are trained to leave out their personal preferences and to evaluate products using specific criteria. Added to this, a small trained panel (usually $n \cong 10$) could never be representative of a target market (Stone & Sidel, 2004). Thus, hedonic perception of products by a few trained assessors does not represent naïve consumers' wide and varied perception and cannot be regarded as a measure of the potential performance of the product in the marketplace (Lawless & Heymann, 2010; O'Mahony, 1979).

Conversely, consumers have been traditionally regarded as not capable of performing analytical tasks and evaluating the sensory characteristics of products in a reliable way (Meilgaard et al., 1999). According to Stone & Sidel (2004), conducting analytical tests with untrained assessors poses several risks to the validity of the results, which are usually underappreciated. However, two decades ago Moskowitz (1996) challenged this idea and claimed that consumers were actually able to accurately rate the intensity of the sensory characteristics of products, providing similar results to trained assessors' panels. Moskowitz' article was strongly criticized (Dugle, 1997; Hough, 1998), and initiated a strong debate in the sensory and consumer field that has been ongoing until now. The development of various alternative flexible methods for sensory characterization, which can be adapted to panels with different degrees of training (Liu, Schou Grønbeck, Di Monaco, Giacalone, & Bredie, 2016), has fueled and agitated the debate about the use of untrained assessors for analytical tasks in sensory science (Valentin, Chollet, Lelièvre, & Abdi, 2012; Varela & Ares, 2012). In the last decade, the line between trained and consumer panels for analytical tests has blurred and is expected to continue to do so (Meiselman, 2013).

In this context, the aim of the present opinion paper is to discuss some of the most relevant issues that have been involved in the discussion of whether consumer or trained assessor panels are appropriate for a specific application and to provide recommendations for practitioners on this respect.

2. Revisiting the arguments against the use of consumers for analytical tasks

Trained assessors have been a cornerstone of sensory evaluation since its establishment as a scientific discipline, which can be probably traced down to the use of professional tasters or experts that worked in the food, beverage and personal care industries since the beginning of the 20th century (Meilgaard et al., 1999). Assessor selection and training have been considered one of the basis of the objectivity and validity of sensory data, as trained assessors have been regarded as instruments that record what they perceive with their senses.

Trained assessor panels have been strongly recommended to provide actionable information in new product development and quality control, as well as to fully characterize the sensory properties of food and non-food products. The use of trained assessors over consumers to perform analytical tasks has been justified based on three main arguments: sensory acuity, reliability and cost efficiency (Moskowitz, 1996). Added to this, consumers have been highlighted to act in a "non-analytic frame of mind" and to not have enough knowledge about specific attributes, confusing some of them (Lawless & Heymann, 2010). In the following sections these arguments are discussed in the light of results from recent scientific studies, as well as methodological and practical considerations, including actual common practices in industrial and academic environments.

2.1. Sensory acuity or familiarity with experimental procedures?

Trained assessors are selected based on their sensory acuity (Lawless & Heymann, 2010), meaning that, on average, they are expected to be more sensitive than naïve consumers. According to Stone & Sidel (2004), 30% of the people who usually volunteer to participate in a panel do not meet the qualifying criteria because they do not reach the minimum level of sensitivity and reliability. This simple and basic step in their selection implies that trained assessor panels may be more sensitive than consumers for identifying specific sensory characteristics or detecting differences between samples. However, although trained assessors usually outperform consumers in their perceptual and verbal abilities for sample evaluation, it is not always the case.

Several studies have shown that training improves assessors' ability to discriminate among samples (Cardello et al., 1982; Clapperton & Piggott, 1979; Fernández-Vázquez, Stinco, Hernanz, Heredia, Vicario, 2013; Guerrero, Gou & Arnau, 1997; Ishii, Kawaguchi, O'Mahony, & Rousseau, 2007; Labbe, Rytz, & Hugi, 2004; Sawyer, Cardello, & Prell, 1988; Solomon, 1990). Similarly, Peron & Allen (1988) reported that perceptual training increases assessors' ability to detect beer specific flavours, whereas Cain (1979) showed that practice and feedback improved people's ability to identify odours.

However, a large number of studies have shown no superiority of trained assessors over consumers. Several studies have shown no effect of training on discrimination (Roberts & Vickers; 1994, Wolters & Allchurch, 1994; Chambers & Smith; 1993). Similarly, olfactory thresholds have been reported to not differ between trained and untrained assessors (Bende & Nordin, 1997; Parr, Heatherbell, & White, 2002). Besides, according to Lawless (1984) the difference between experts and novices in their ability to describe white wine is small.

Differences between trained assessors and consumers are mainly found on stimuli on which the former have been previously trained. According to Chollet, Valentin, & Abdi (2005) trained assessors do not generalize their perceptual learning and, consequently, they do not differ from consumers in their ability to discriminate unknown stimuli. These authors explained this lack of perceptual transfer to perceptual learning: assessors learn to extract and encode the sensory characteristics that are optimal for discriminating a set of samples, which may not be useful to discriminate among other stimuli.

Therefore, although it has been widely accepted that trained assessors outperform consumers, their superiority seems to be mainly related to their familiarity with the experimental procedures used for sample evaluation (Ishii et al., 2007), as well as their ability to describe their perception (Chollet & Valentin, 2001). In this sense, recent studies have shown short familiarization steps can improve consumer performance in analytical tests (Liu et al., 2016; Jaeger et al., 2017).

Even if trained assessors are more discriminative than consumers, the main question is whether this matters. Do we want to base our decisions on the perception of assessors highly trained in detecting small differences among samples? The answer to this question is "It depends". When the aim of the study is to assure that sensory differences between products are negligible for consumers, trained assessors may provide conservative responses for project managers. However, when trained assessors are able to detect differences among samples, the key question is whether the difference between products is relevant for consumers. In these situations, discrimination-testing programs conducted with trained assessors require tools relating the discriminative ability of trained and consumer panels

147 (Rousseau, 2015). Therefore, consumer panels are indeed relevant for decision making to
148 determine when the sensory differences perceived by trained assessors translate into sensory
149 or hedonic differences for consumers.
150
151 **2.2. Reliability: A matter of adequacy of experimental procedures**
152 Another relevant argument against the use of consumers for analytical tasks has been
153 related to the fact that consumer attribute information is not reliable because they face several
154 difficulties for understanding product attributes and scales (Muñoz, 1997; Stone & Sidel,
155 2004). However, this direct comparison is not fair, as trained assessors use a common and
156 standardized vocabulary, previously learnt evaluation protocols, and are thoroughly trained to
157 rate the intensity of sensory attributes using scales with clearly defined references (Lawless
158 & Heymann, 2010). On the other hand, when consumers are asked to evaluate specific
159 sensory attributes they are not usually given precise instructions about how to evaluate or rate
160 the products. In this sense, it should be taken into account that a limited amount of training
161 can largely improve assessor performance in analytical tasks (Liu et al., 2016; Jaeger et al.,
162 2017; Saint Saint-Eve, Lenfant, Teillet, Pineau, & Martin, 2011). Similarly, for descriptive
163 analysis it has been reported that the first few sessions provide the biggest gains in terms of
164 ability to discriminate among samples and increasing consensus among assessors (Byrne,
165 Bredie, & Martens, 1999; Byrne, O'Sullivan, Dijksterhuis, Bredie, & Martens, 2001).
166 Consumer interpretation of specific sensory attributes may be highly heterogeneous
167 as they may have different interpretation of the meaning of specific sensory attributes. This
168 has been previously shown for complex texture attributes such as creaminess (Antmann,
169 Ares, Varela, Salvador, Coste, & Fiszman, 2011). Lack of consensus in consumer evaluations
170 of attribute intensities using scales is also expected, as consumers might be strongly
171 influenced by their personal preferences and previous experiences with the product category.
172 Ares, Bruzzone, & Giménez (2011) reported large heterogeneity in consumer intensity ratings
173 of texture attributes (particularly for complex attributes, such as creaminess and
174 homogeneous) and showed that the great majority of consumers were not able to use

unstructured intensity scales to indicate differences in the texture of a set of vanilla milk desserts. However, at the average level consumers provided the same information than trained assessors regarding significant differences among samples, despite differences in the range of the scale used for sample evaluation. Similar results have been reported by Bruzzone, Vidal, Antúnez, Giménez, Deliza, & Ares (2015), Husson, Le Dien & Pagès (2011), Moskowitz (1996), and Worch, Lê, & Punter (2010).

Although average intensity scores from consumers have been shown to be similar to those obtained with trained assessors in several specific studies, care must be taken when interpreting intensity ratings from consumers as they do not have common references for scaling. The use of intensity scales for sample evaluation is basically an extension of the experimental procedures used with trained assessors. In the authors' opinion, experimental procedures should be adapted to the characteristics of the assessors involved in the test. Therefore, when sensory characterization with consumers is sought, researchers are encouraged to use standardization procedures to remove individual differences in scale use or to apply methodologies that get rid of individual differences in scaling.

For example, methodologies based on ranking (e.g. flash profile), attribute selection (e.g. check-all-that-apply questions) or global similarities and differences among samples (e.g. sorting or projective mapping) can be a better choice for sensory characterization with consumers than scales. In this sense, research has shown that the former methodologies provide reliable results and that in most instances provide comparable results to descriptive analysis with trained assessors (Ares et al., 2015; Chollet, Lelièvre, Abdi, & Valentin, 2011; Delarue & Sieffermann, 2004; Moussaoui & Varela, 2010; Risvik, McEwan, & Rodbotten, 1997). Besides, consumer panels have been shown to be repeatable at the aggregate level (e.g. Jaeger et al., 2013; Vidal et al., 2014; Vidal, Jaeger, Antúnez, Giménez, & Ares, 2016). However, tools for evaluating the reliability of consumer panels are still necessary. Researchers should be able to demonstrate the reliability of their data collected with consumer panels as they usually do with trained assessors (Ares, 2015).

Regarding sample description, it should be taken into account that trained assessors

tend to have a more precise vocabulary than consumers and to use it more efficiently to describe samples (Chollet & Valentin, 2001). Consumers usually use less technical, more ambiguous and redundant terms, as well as words related to hedonics or attribute intensity to describe samples than trained assessors (Moskowitz et al., 2003; Lelièvre, Chollet, Valentin, & Abdi, 2008; Veramendi, Herencia, & Ares, 2013). Although this may be seen as a disadvantage, it is important to stress that it may not be a problem when the objective of the study is to discriminate among samples. Besides, working with consumer vocabulary enables the identification of relevant terms for the design marketing and communication campaigns.

Added to the perceptual aptitude itself, a good sensory panelist is not only expected to be more sensitive than the average, but also to be articulate and to have a good descriptive ability. Besides, as concept formation is dependent on prior experience, when assessors are trained for descriptive analysis, they are taught how to create their own scientific language for the product category of interest, creating a "frame of reference" for the panel as a group (Murray, 2001; Lawless & Heymann, 2010). So, in a way, panelists are first selected to be articulate, being able to express their perception, and subsequent training makes them able to describe products in a homogenous way. Consumers, on the contrary, could generate long lists of words, much less consensual – and sometimes quite complex to interpret – but undoubtedly richer. Consumer vocabulary expands the possibilities of capturing consumers' sensory perceptions in their own words, as it has been shown in many studies that have compared methods of sensory description with consumers (Delarue, 2015; Fiszman, Salgado, Orrego, & Ares, 2015; Moussaoui & Varela, 2010; Veinand et al., 2011; Varela & Ares, 2012, Valentin at al., 2012).

In summary, it has been demonstrated that consumers are able to reliably evaluate the sensory characteristics of products, even if large individual differences in how they describe products and rate the intensity of sensory attributes exist. Researchers are encouraged to use methodologies adapted that take into account these differences as well as the lack of training.

2.3. Cost efficiency: A matter of objective and context

Trained assessor panels have been regarded as a cost efficient option as they usually involve a limited number of people that work at the test location. However, it should be taken into account that creating and maintaining a well-trained panel can be expensive in several circumstances. For this reason, the relative cost of trained and consumer panels strongly depends on the objective of the study and context.

In the authors' experience, several big companies need sensory information for the development of a specific product a few times a year, which makes consumer panels the most cost-efficient option. Also, several small food companies usually cannot afford to maintain a trained panel and therefore consumer panels consist of the only alternative to gather objective information for decision making.

On the contrary, when sensory information is needed on a daily or even monthly basis, trained panels continue to be the most cost-efficient option. Nevertheless, when companies are already doing consumer testing for new product development, the use of alternative methods for sensory characterization can give them many interesting inputs without the need of having a trained panel.

Therefore, the cost efficiency of trained assessor and consumers for analytical testing depends on the aim of the study. Researchers should analyze the costs associated with each panel for each specific project.

3. Additional arguments regarding the use of trained and consumer panels

Apart from the traditional arguments involved in the discussion of whether consumer panels should be used for analytical testing, there are several additional issues that should be taken into account. The following sections address some of the issues that in the authors' view have not received enough attention yet.

3.1. Can trained assessors be considered as analytical instruments?

Trained assessors have been traditionally regarded as analytical instruments, capable

of providing accurate and repeatable evaluations of the sensory characteristics of products. But, are human beings really able to behave as analytical instruments? The answer is no. Sensory perception does not only depend on the physicochemical characteristics of products. Instead, it depends on several integrated physiological, psychological and physical processes that occur in our brain (Schifferstein, 1996). Frijters (1993) discusses three processes involved from perception of a physical stimulus to an intensity rating: i) transformation of the physical stimulus into a sensation, ii) representation of the stimulus into an internal subjective continuum and storage into working memory, and iii) transformation of the subjective continuum into a response to the experimental task. These processes are influenced by the experimental procedure, the experimental design, changes in physiological or cognitive parameters during the test and contextual information about the stimulus (Schifferstein, 1996). Therefore, responses from trained assessors to any analytical tests should be considered as context-dependent and not as absolute responses from an analytic instrument.

Furthermore, even if trained assessors could behave as analytical instruments, their data would only serve for limited purposes as they would not reflect what consumers perceive or how they behave in their daily life. In his nice paper, Köster (2003) discusses several fallacies that are usually encountered in sensory and consumer science. In the following sub-sections, the implications of some of the fallacies highlighted by Köster in the discussion of whether consumers or trained panels should be used for analytical tests.

3.2. Much more than sensory acuity

As discussed above, sensory perception is not only a question of sensitivity; attention and cognitive processing of the signals we attend to are also important variables in this discussion. Perceptual attention seems to determine what we consciously perceive- and subsequently describe. We only perceive that to which we attend to, although many times we perceive much more than we seem to notice (Noë & O'Regan, 2000). In particular, these two phenomena would compete when assessors are acting in analytical –focusing on particular individual attributes - vs holistic mode. Some researchers in the area have suggested that the

process of synthesis (the way sensory information about products is analyzed and processed) might be different between consumers and trained panelists (Jaeger et al., 2000), and even within the same descriptive panel because of the different cognitive styles (Varela et al. 2014; Vidal et al., 2015; Antúnez et al., 2015). Further than this, individual differences in preferred ways of processing information or cognitive styles are also expected to influence responses to analytical tasks. In particular, the wholistic-analytic dimension, which separates people who have tendency to process information globally (wholistic), and those who have tendency to process information in detail and to focus on specific characteristics (analytic) (Peterson & Deary, 2006), could be very much related to the different performance of individual assessors within a trained panel. However, this would also mean that some consumers, even if less sensitive, could be more analytically framed and might perform better in analytical tasks. Kinner and Bongartz (2015) also suggested the idea of the difference between distinct cognitive reflection types (slow and fast thinkers) and their ability to discriminate in consumer tests. Their results showed that that slow thinkers had a higher ability to discriminate between samples in consumer testing, but this could also well be the case in sensory testing. This is a completely new area, which remains to be explored.

Vocabulary generation and training in classic descriptive analysis aims at generating a list of measurable attributes or scorecard (Stone & Siedel, 2004; Stone, 2015). However, what happens when a particular attribute in a product set is not easily measurable? Possible cases are when the particular attribute is at the same level in all the products of the category under study, or when it is present in a low, just noticeable intensity. Many times, those attributes can be disregarded by trained panels, taken out of the scorecard because they do not discriminate among samples. However, those attributes might be in fact drivers of consumer liking or disliking. Those particular attributes could be an off-note, or a positive “must have” attribute, even if present in low intensity. Sometimes attributes with high intensities might be not discriminative for the trained panel, but be determinant of consumer acceptance or rejection, for instance because of an unbalance caused by the levels of other attributes. Let’s take the example of espresso coffee. Espresso brewed in different machines or with

315 different brewing parameters can have big variations in the amount and characteristics of
316 crema (bubble size, viscosity, etc.), so you can brew two cups using exactly the same coffee,
317 resulting in completely different consumer experiences. Those two coffees can have no
318 significant differences in bitterness intensity rating as assessed by a highly trained panel, but
319 bitterness will be perceived by consumers at completely different levels because of the
320 mouthfeel effect generated by the crema. Consumers could reject one of the samples because
321 of its enhanced bitterness, and they could easily describe their perception as: “this coffee is
322 too bitter, I don’t like it”. Even if mouthfeel could in principle also affect the perception of the
323 panel, highly trained analytical assessors are usually able to “deconstruct” the sensory profile
324 and to assess the individual attributes independently. Added to this, many times when tasting
325 beverages in individual servings, particularly when focusing on flavor, samples may be bulked
326 in thermoses before being tasted by the panel (to account for machine differences, to get
327 homogenous samples among the panel and control temperature throughout the tasting
328 session). For the case of the example, following these kind of procedures the effect of crema
329 would be lost for the trained panel.

330 In several circumstances, consumers could in fact be even a better sensory tool than
331 trained assessors, because of their particular cognitive thinking styles or their language
332 capabilities, or because they focus more on the characteristics that drive their preferences.
333 This brings us to the topic of ecological validity of the tasting, which will be discussed in the
334 next section.

335

336 **3.3. Ecological validity of analytical measurements**

337 One of the outcomes of the final panel discussion of the 2015 Pangborn Sensory
338 Science Symposium highlighted the need to increase the ecological validity of both sensory
339 and consumer science measurements, and, particularly, to account for individual differences
340 in perception and decision making (Jaeger et al., 2016). This is very important when thinking
341 about preferences, but not less important for food perception and description, when the aim
342 is to explain and predict consumer preferences.

Sample preparation is the first issue one can think about in this sense. In an analytical test with trained panels, the samples are often prepared in a way that minimize sample variation in order to avoid adding another source of variability to the data. This include, among many others, practices such as: bulking of beverages, sample cutting to homogenize sample sizes, taking out the crust of bread products, chocolate melting and re-forming to get rid of brands or recognizable shapes, cutting bite-size pieces or serving semi-solid samples directly as a spoonful to assess temporal perception, or using of red-light to avoid colour influence on flavor perception. These practices will obviously make the panel assessment quite far to what consumers will experience in real life consumption.

Going to the sensory perception itself, consumers usually spend little time and do not often engage in deep cognitive processing to evaluate the characteristics of food products when making their food choices (van't Riet et al., 2011). Nevertheless, when performing classical analytical testing, trained panelists are encouraged to engage in deep analytical processing, which is also often the case in some consumer based descriptive tests, which would not reflect how consumers process information when choosing or consuming food in their everyday life.

For example, classical discrimination tests, such as paired comparison and triangle tests, lack ecological validity as consumers would very unlikely evaluate two products from different batches at the same time. In this sense, the A not-A methods provide a more ecologically valid evaluation. Assessors are familiarized with a product and are then given and are asked to indicate whether they are identical to the first product or not (Lee, van Hout, & O'Mahony, 2007). This type of evaluation is more similar what they would do in their real life when comparing to batches of the same product: they would have to compare the batch they are consuming with their memory of the previous consumed batch. Recent research has shown that the A not-A test can be superior in discrimination than the triangle or tetrad test (Jeong, Kang, Jeong, Song, Hautus, & Lee, 2016).

Something similar happens with descriptive methods, some methods induce an analytical evaluation, focusing on specific individual attributes, whereas other methods enable

371 a more holistic evaluation based on products as a whole (Sloman, 1996). This could be the
372 case of Free Sorting, Projective Mapping, or even Polarized Sensory Positioning (PSP) and
373 Pivot Profile (Varela & Ares, 2012). Those methods are most of the times used with consumers
374 or semi-trained panels, but could also be used with trained sensory panels. The issue though,
375 could be that highly trained panelists are not always comfortable when using holistic
376 approaches to sensory description, if they do not frequently use them within their method
377 portfolio. In this case, a slightly more analytical approach as PSP could be a good middle-way
378 solution.

379 In terms of ecological validity in a wider concept, the importance of context on sensory
380 perception has been probably underestimated as analytical tests are usually conducted blind,
381 without any type of contextual information. However, the expectations generated by packages,
382 labels, or even prior information have been reported to extensively influence how people
383 perceive products (Cardello, 2007; Piqueras-Fiszman & Spence, 2015). In addition, the
384 processes involved in the transformation of a sensory stimulus into an intensity rating have
385 been reported to be influenced by contextual information (Schifferstein, 1996). This suggests
386 that results from analytical tests are expected to be influenced by context and external
387 information about products. However, this area of research has not received enough attention
388 yet and could contribute to a better understanding of how expectations shape sensory
389 perception. In the future, one could think of performing analytical tests in a natural situation,
390 immersive reality or evoked contexts in order to consider the situational and contextual factors
391 that influence sensory perception (Jaeger et al., 2016). This has been used with success in
392 affective tests in the last years and might as well be relevant to obtain more ecologically valid
393 analytical data in the future.

394 Further than this, trained panels do not usually take into account individual differences
395 in sensory perception. Interest in understanding how individual differences on sensory
396 perception (PTC, PROP, Thermal Taster Status, or other pheno- and genotypic differences)
397 influence consumer hedonic reaction to food products and their food choices is expected to
398 increase in the future (Jaeger et al., 2016). This type of research should be carried out with a

large number of participants in order to account for those differences, which is not normally the case with trained sensory panels. In this context, consumer panels will be of great importance. This could be an important factor to have in mind in the future, particularly when thinking of food companies wanting to develop personalized products.

Again, consumer panels seem to be relevant sensory descriptive tools when highlighting the sensory characteristics that underlie hedonic perception, when describing their own perception and when more ecologically valid tests are sought.

3.4. On not-that-representative consumers and not-that-trained trained panelists

When discussing the use of trained and consumer panels it is worth highlighting the importance of best practices in the design of analytical tests. Sometimes when performing a sensory or a consumer test, objectives are discussed, methodological implications evaluated and decided, and then, reality bites: consumers are not-that-representative, trained panelists are not that-trained, and sometimes even the trained panel is actually not-that-panel. These situations frequently happen in both academic and industrial research settings.

In many academic research papers, we can find numerous examples of consumer panels that are actually “student panels”, very limited consumer panels in terms of number of participants, or a not representative or relevant population for answering the research question under study. This could be quite relevant when drawing conclusions on preference or food choice, but it could of course be also relevant when exploring product profiling as conclusions are drawn in terms of the perception of a particular population. Apart from the representativeness of a consumer panel there is also the reliability issue. There are some recommendations in terms of minimum number or panelists for alternative product profiling techniques with consumers like CATA and Projective Mapping to ensure the stability of the obtained configurations (Vidal et al., 2014; Ares et al., 2014). However, best practices are not always followed. The issue of small, not representative consumer panels is also frequent in industrial R&D settings, mostly for limited resources allocated. Many big companies make use of their internal employees to run acceptability tests and more recently have started to gather

sensory data concurrently (quite often CATA). The main danger here is that preference data are most probably biased. However, sensory data collected in those tests could also be compromised, as per the same comments above. In a recent study, Cardinal et al. (2015) highlighted a consumer segment effect when comparing acceptability ratings and responses to CATA questions collected with target consumers versus convenience consumer samples (food science related consumers), which can lead to erroneous product development directions. Thus, recruitment of users of the category is not only relevant when collecting data (Lawless & Heymann, 2010), but also for sensory profiling objectives.

Online consumer panels are also worth mentioning here. With the widespread of internet and social media, it is quite simple to put together a survey and reach consumers with a link in an e-mailing list, a Facebook page or a tweet. With regards to analytical tests, one could think of profiling food concepts, labels or packaging, for example. The use of online tools for this could be tempting and indeed useful if it is possible to know the source of the data, but in the same way very risky if we do not get a clear view of whom these consumers are, with a result of a potentially big, but unrepresentative panel. On the other hand, sources like Facebook fan pages or specialist blogs could be a great source of direct information form likers and heavy users of the products, which could be advantageous if feedback is wanted from heavy users.

Regarding trained panels, the authors have frequently seen cases, particularly in industrial settings, in which decisions are made based on results from poorly trained and maintained panels. It is common practice to use internal employees that, even if quite unbiased and recruited from outside of the product development teams, are not very steady in terms of participation in the panel, as these activities quite often come last in their to-do lists. In fact, this produces a “pool of semi-trained assessors” rather than a trained panel. Moreover, even when the panel is more or less constant as a group, many times the training opportunities are scarce, and their performance consequently poor.

Particular mention should also be made to “expert panels”, used in industries such as coffee, perfume, tea, tobacco or wine. These tasters are usually very sensitive to many

455 characteristics of a single product through experience and are able to make rapid judgements
456 for sample and material selection. They are usually not selected or trained, and work
457 individually or in small groups, but not as part of a calibrated panel. Many times, they also
458 know in advance certain information about the products. Feria-Morales (2002) does a good
459 account of the flaws and biases of using expert panels in the coffee industry, recommending
460 the shift towards the use of standard sensory procedures and trained sensory panels. Zamora
461 & Guirao (2002) compared trained assessors with experts for wine assessment, concluding
462 that the trained panel reached a higher level of consensus, while the experts were more
463 discriminative among attributes. Lawless and Heymann (2010) nicely explain the historical
464 bases of expert panels and highlight that for quality assessment of certain food commodities
465 such as olive oil, they could still have a place in the sensory toolbox, guided by very precise
466 written standards of the International Olive Oil Council (COI), for example. However, these
467 methods are not well suited to formulated or more complex foods that do not fall into the
468 category of a standardized commodity.

469 Thus, apart from considering the objectives of a test, one should do a reality check. Is
470 my trained panel really a trained panel? Is it worth spending time and effort to collect data with
471 the “trained panel” and get a not very reliable outcome? Or shall I explore analytical tests with
472 consumers instead or make use of methods better suited for semi-trained assessors? When
473 working with consumers, one should also look at representativeness including frequency of
474 usage of the product, and best practices leading to validity and reliability of the obtained
475 results.

476

477 **4. Recommendations for the use of consumer panels for analytical tasks**

478 Research in the last decades has shown that consumer panels are indeed able to
479 evaluate the sensory characteristics of products and to provide similar results to trained
480 assessors. However, experimental procedures for collecting analytical tasks with consumers
481 cannot be identical to those used with trained assessors as they should take into account the
482 lack of training.

Although both panels can provide reliable results, the answer to the "trained assessors vs. consumers" controversy strongly depends on the objective of the study. In specific circumstances, trained panels are clearly the best alternative because untrained consumer panels are not feasible and/or would not provide reliable results.

Quality control is the best example of a specific task in which trained panels could probably never be replaced by consumer panels. In quality control, trained assessors are needed to detect small variations in the product and to detect the presence of sensory defects before a batch goes out to the market (Moskowitz, 1997). Research has shown that some consumers may not be able to detect sensory defects (e.g. Mörlein, 2012), or even prefer defective samples (e.g. Ramírez, Hough, & Contarini, 2007). Mismatch between quality ratings given by experts and consumer hedonic scores has been reported to exist, particularly in complex products such as wine and olive oil. In this last product category, Delgado & Guinard (2011) showed that for the majority of consumers hedonic scores did not match quality experts' ratings as defects, such as fusty, musty and rancid, were identified as drivers of liking. Consumers may not be able to detect off-flavours or to associate them with product deterioration, suggesting that trained assessor panels may be always preferred to consumer panels for this type of task. Besides, even if consumers could accurately detect and identify sensory defects it would not be feasible to repeatedly gather consumers to evaluate all the batches produced by a company.

On the other hand, if sensory information is going to be used to guide product development or to identify drivers of consumers' liking, trained and consumer panels most of the time provide similar information (e.g. Bruzzone et al., 2015) and therefore consumer panels tend to be a good methodological choice. This is particularly the case in the first stages of new product development, as prototypes can be selected based on results from consumer panels using alternative methodologies. However, it should be taken into account that when dealing with subtle differences among samples, trained assessors are expected to outperform consumers in their ability to discriminate among samples (Antúnez et al., 2016; Ares et al., 2015; Torri et al., 2013). In addition, it should be acknowledged that trained assessor data

may be more actionable than consumer responses in new product development (Moskowitz et al., 2003). Although consumers can accurately detect differences among samples, it may be difficult to translate consumer data to actionable directions to product developers, particularly during product reformulation. Trained panels usually provide accurate intensity information that enables product developer to make specific changes in product formulation to achieve the desirable modification in the sensory characteristics of products. This type of information would be difficult to obtain with consumer panels. Besides, given the iterative nature of new product development, it may be necessary to compare prototypes obtained in different moments in time. In these situations, it may be difficult to compare results obtained with consumer panels, although methodologies based on the comparison with references can provide accurate results (Antúnez, Salvador, de Saldamando, Varela, Giménez, & Ares, 2015; Teillet, Schlich, Urbano, Cordelle, & Guichard, 2010). A similar limitation may be faced when evaluating very complex or saturating products.

The ecological validity of analytical measurements should also be taken into account as it can largely affect the ability to predict consumer hedonic perception and choice, regardless of the type of panel being considered. Researchers are encouraged to further study the influence of contextual and situational variables on sensory perception and results from analytical tests.

5. Conclusions and remaining challenges

The debate of whether consumer or trained panels should conduct analytical tests has already come to an end as the hypothesis that consumers are capable of evaluating the sensory characteristics of products has become increasingly accepted within the sensory science community. Research conducted during the last decade has shown that, using appropriate methodologies, consumers are able to provide accurate and reliable information about the sensory characteristics of products. According to the authors, whether consumers or trained assessors should be used depends on the specific circumstances of the study. Objectives and resources must be carefully considered, together with the ecological validity

implications around the specific research questions of the project. In most situations, consumers can replace trained assessors and provide actionable information to guide decision making in both industrial and academic applications. However, sensory and consumer researchers should be aware that trained assessors are still necessary in several specific situations. We hope that the issues raised in the present paper could shed light on which situations each panel can be used, contributing to the definition of new best practices in the field. In addition, it seems that the time has come for sensory science professors to update the curricula of their courses to introduce their students to the current views about analytical tests and put away the consumer vs. trained assessor dichotomy.

Acknowledgements

The authors are thankful for the financial support received from Comisión Sectorial de Investigación Científica (Universidad de la República, Uruguay), the Norwegian Foundation for Research Levy on Agricultural Products through the research program "Sensory strategies and consumer insight for healthy and palatable food" and to FFL and the Research Council of Norway through the RapidCheck project. The authors are also grateful to Herb Meiselman for his invitation to write this opinion paper.

References

Antmann, G., Ares, G., Varela, P., Salvador, A., Coste, B., Fiszman, S.M. (2011). Consumers' creaminess concept perception: A cross-cultural study in three speaking countries. *Journal of Sensory Studies*, 42, 50-60.

Antúnez, A., Salvador, A., de Saldamando, L., Varela, P., Giménez, A., Ares, G. (2015). Evaluation of data aggregation in Polarized Sensory Positioning. *Journal of Sensory Studies*, 30, 46-55.

566 Antúnez, L., Oliveira, D., Vidal, L., Ares, G., Næs, T., Varela, P. (2015) Influence of
567 wholistic/analytic cognitive style on results from projective mapping. Poster presentation
568 at the 11th Pangborn Sensory Science Symposium, 23–27 August 2015, Gothenburg,
569 Sweden
570 Antúnez, L., Vidal, L., de Saldamando, L., Giménez, A., Ares, G. (2016). Comparison of
571 consumer-based methodologies for sensory characterization: Case study with four
572 sample sets of powdered drinks. *Food Quality and Preference*, In press, doi:
573 http://dx.doi.org/10.1016/j.foodqual.2016.09.013
574 Ares, G. (2015). Methodological challenges in sensory characterization. *Current Opinion in*
575 *Food Science*, 3, 1-5.
576 Ares, G., Antúnez, L., Bruzzone, F., Vidal, L., Giménez, A., Pineau, B., et al. (2015).
577 Comparison of sensory product profiles generated by trained assessors and consumers
578 using CATA questions: Four case studies with complex and/or similar samples. *Food*
579 *Quality and Preference*, 45, 75-86.
580 Ares, G., Bruzzone, F., & Giménez, A. (2011). Is a consumer panel able to reliably evaluate
581 the texture of dairy desserts using unstructured intensity scales? Evaluation of global and
582 individual performance. *Journal of Sensory Studies*, 26, 363-370.
583 Ares, G., Tárrega, A., Izquierdo, L. Jaeger, S. (2014) Investigation of the number of consumers
584 necessary to obtain stable sample and descriptor configurations from check-all-that-apply
585 (CATA) questions. *Food Quality and Preference*, 31, 135–141
586 Bende, M., & Nordin, S. (1997). Perceptual learning in olfaction: professional wine tasters
587 versus controls. *Physiology and Behavior*, 62, 1065–1070.
588 Bruzzone, F., Ares, G., & Giménez, A. (2012). Consumers' texture perception of milk desserts
589 II—Comparison with trained assessors' data. *Journal of Texture Studies*, 43, 214–226.
590 Bruzzone, F., Vidal, L., Antúnez, L., Giménez, A., Deliza, R., & Ares, G. (2015). Comparison
591 of intensity scales and CATA questions in new product development: Sensory
592 characterisation and directions for product reformulation of milk desserts. *Food Quality*
593 *and Preference,* 44, 183-193.

Byrne, D. V., Bredie, W. L. P., & Martens, M. (1999). Development of a sensory vocabulary for warmed-over flavour. Part II: In chicken meat. *Journal of Sensory Studies,* 14, 67–78.

Byrne, D. V., O'Sullivan, M. G., Dijksterhuis, G. B., Bredie, W. L. P., & Martens, M. (2001). Sensory panel consistency during development of a vocabulary for warmed-over flavour. *Food Quality and Preference*, 12, 171–187.

Cain, W. S. (1979). To know with the nose: Keys to odor identification. *Science*, 203, 467–470.

Cardello, A.V. (2007). Measuring consumer expectations to improve food product development. In H.J.H. MacFie (Ed.), *Consumer-led food product development*, pp. 223–261. Cambridge: Woodhead Publishing

Cardello, A.V., Maller, O., Kapsalis, J.G., Segars, R.A., Sawyer, F.M., Murphy, C., & Moskowitz, H. (1992). Perception of texture by trained and consumer panelists. *Journal of Food Science*, 47, 1186-1197.

Cardinal, P., Zamora, M. C., Chambers, E., Carbonell Barrachina, Á. and Hough, G. (2015), Convenience Sampling for Acceptability and CATA Measurements May Provide Inaccurate Results: A Case Study with Fruit-Flavored Powdered beverages Tested in Argentina, Spain and U.S.A.. *J Sens Stud,* 30: 295–304

Chambers, E., IV, & Smith, E. A. (1993). Effects of testing experience on performance of trained sensory panelists. *Journal of Sensory Studies*, 8, 155–166.

Chollet, S., Lelièvre, M., Abdi, H., & Valentin, D. (2011). Sort and beer: Everything you wanted to know about the sorting task but did not dare to ask. *Food Quality and Preference,* 22, 507–520.

Chollet, S., & Valentin, D. (2001). Impact of training on beer flavor perception and description Are trained and untrained subjects really different? *Journal of Sensory Studies*, 16, 601-618.

Chollet, S., Valentin, D., & Abdi, H. (2005). Do trained assessors generalize their knowledge to new stimuli? *Food Quality and Preference,* 16, 13-23

621 Clapperton, J. F., & Piggott, J. R. (1979). Flavor characterization by trained and untrained
622 assessors. *Journal of Institute of Brewing*, 85, 275–277.
623 Delarue, J. and Sieffermann, J.-M. (2004). Sensory mapping using Flash profile. Comparison
624 with a conventional descriptive method for the evaluation of the flavour of fruit dairy
625 products*. Food Quality and Preference,* 15, 383–392.
626 Delarue, J. (2015) The use of rapid sensory methods in R&D and research: an introduction.
627 In: Rapid Sensory Profiling Techniques and Related Methods Applications in New Product
628 Development and Consumer Research. Edited by Delarue, J. Lawlor, JB. and Rogeaux,
629 M.. Woodhead Publishing, Cambridge. ISBN 978-1-78242-248-8
630 Delgado, C., & Guinard, J. -X. (2011). How do consumer hedonic ratings for extra virgin olive
631 oil relate to quality ratings by experts and descriptive analysis ratings? *Food Quality and*
632 *Preference*, 22(2), 213–225.
633 Dugle, J. (1997). Note on ‘‘experts versus consumers: a comparison’’. *Journal of Sensory*
634 *Studies*, 12, 147–153.
635 Feria-Morales, A. M. (2002) Examining the case of green coffee to illustrate the limitations of
636 grading systems/expert tasters in sensory evaluation for quality control. *Food Quality and*
637 *Preference, 1*3: 355-367
638 Fernández-Vázquez, R, Stinco, C.M., Hernanz, D., Heredia, F.J., & Vicario, I.M. (2013).
639 Colour training and colour differences thresholds in orange juice. *Food Quality and*
640 *Preference*, 30, 320-327.
641 Fiszman, S.M., Salgado, N.., Orrego, C.E., & Ares, G. (2015). Comparison of methods for
642 generating sensory vocabulary with consumers: A case study with two types of satiating
643 foods. *Food Quality and Preference*, 44, 111-118.
644 Frijters, J.E.R. (1993). Functional measurement in the study of mixture percepts. *Chemical*
645 *Senses*, 18, 93-100.
646 Guerrero, L., Gou, P. & Arnau, J. (1997). Descriptive analysis of toasted almonds: A
647 comparison between expert and semitrained assessors. *Journal of Sensory Studies*, 12,
648 39–54.

649 Hough, G. (1998). Experts versus consumers: a critique. *Journal of Sensory Studies*, 13, 285-
650 289.
651 Husson, F., Le Dien, S., & Pagès, J. (2001). Which value can be granted to sensory profiles
652 given by consumers? Methodology and results. *Food Quality and Preference*, 12, 291–
653 296.
654 Ishii, R., Kawaguchi, H., O'Mahony, M., & Rousseau, B. (2007). Relating consumer and
655 trained panels' discriminative sensitivities using vanilla flavoured ice cream as medium.
656 *Food Quality and Preference*, 18, 89-96.
657 Jaeger, S.R., Beresford, M.K., Hunter, D.C., Alcaire, F., Castura, J.C., Ares, G. (2017). Does
658 a familiarization step influence results from a TCATA task? *Food Quality and Preference,*
659 55, 91-97.
660 Jaeger, S. R., Chheang, S. L., Jin, D., Bava, C. M., Gimenez, A., Vidal, L., et al. (2013). Check-
661 all-that-apply (CATA) responses elicited by consumers: Within-assessor reproducibility
662 and stability of sensory product characterizations. *Food Quality and Preference*, 30, 56–
663 67.
664 Jaeger, S.R., Wakeling, I.N., & MacFie, H.J.H. (2000). Behavioural extensions to preference
665 mapping: the role of synthesis. *Food Quality and Preference*, 11, 349-359.
666 Jaeger, SR., Hort, J., Porcherot, G., Ares, G., Pecore, S., MacFie, HJH. (2016) Future
667 directions in sensory and consumer science: Four perspectives and audience voting.
668 *Food Quality and Preference*, doi:10.1016/j.foodqual.2016.03.006
669 Jeong, Y.N., Kang, B.A., Jeong, M.J., Song, M.J., Hautus, M.J., Lee, H.S. (2016). Sensory
670 discrimination by consumers of multiple stimuli from a reference: Stimulus configuration
671 in A-Not AR and constant-ref. duo-trio superior to triangle and unspecified tetrad? *Food*
672 *Quality and Preference*, 47, 10-22.
673 Kinner, Julia; Bongartz, Annette (2015). Slow or fast thinkers?: Cognitive reflection types and
674 their ability to discriminate in consumer tests. Oral presentation, 11th Pangborn Sensory
675 Science Symposium, Gotheborg, Sweden.

Labbe, D., Rytz, A., & Hugi, A. (2004). Training is a critical step to obtain reliable product profiles in a real food industry context. *Food Quality and Preference*, 15, 341–348.

Lawless, H. T. (1984). Flavor description of white wine by ‘‘expert’’ and nonexpert wine consumers. *Journal of Food Science,* 49, 120–123.

Lawless, H.T., & Heymann, H. (2010). *Sensory Evaluation of Food: Principles and Practices* (Second Edition). New York: Springer.

Lee, H.S., van Hout, D., O'Mahony, M. (2007). Sensory difference tests for margarine: A comparison of R-Indices derived from ranking and A-Not A methods considering response bias and cognitive strategies. Food Quality and Preference, 18, 675-680.

Liu, J., Schou Grønbeck, M., Di Monaco, R., Giacalone, D., Bredie, W. (2016) Performance of Flash Profile and Napping with and without training for describing small sensory differences in a model wine. *Food Quality and Preference*, 48, 41–49.

Köster, E.P. (2003). The psychology of food choice: Some encountered fallacies. *Food Quality and Preference,* 14, 359-373.

Meilgaard, M., Civille, G.V., & Carr, B.T. (1999). *Sensory Evaluation Techniques* (Third Edition). Boca Raton, FL: CRC Press.

Meiselman, H.L. (2013). The future in sensory/consumer research: evolving to a better science. *Food Quality and Preference*, 27, 208-214.

Mörlein, D. (2012). Boar taint: The sensory perspective - Olfactory perception, consumer acceptance and trained sensory panel evaluation of boar taint. *Zuchtungskunde*, 84, 427-438.

Moskowitz, H.R. (1996). Experts versus consumers: A comparison. *Journal of Sensory Studies*, 11, 19–37.

Moskowitz, H.R., Muñoz, A.M., & Gacula, M.C. (2003). *Viewpoints and controversies in sensory science and consumer product testing.* Trumbull, Connecticut: Food & Nutrition Press.

Moussaoui, K. A. & Varela, P. (2010). Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality and Preference*, 21, 1088–1099.

Muñoz, A.M. (1997). Importance, types and applications of consumer data relationships. In A.M. Muñoz (Ed.), *ASTM Manual 30. Relating consumer, descriptive and laboratory data to better understand consumer responses*. West Conshohocken, Penn: ASTM Press.

Murray, J. M., Delahunty, C. M. and Baxter, I. A. (2001). Descriptive analysis: Past, present and future. *Food Research International,* 34, 461–471.

Noë, A. & O'Regan, JK. (2000) Perception, Attention and the Grand Illusion. *Psyche*, 6(15).

O'Mahony, M. 1979. Psychophysical aspects of sensory analysis of dairy products: a critique. *Journal of Dairy Science*, 62, 1954–1962

O'Mahony, M. (1995). Sensory measurement in food science—fitting methods to goals. *Food Technology*, 49, 72–82.

Parr, W. V., Heatherbell, D., & White, K. G. (2002). Demystifying wine expertise: olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. *Chemical Senses,* 27, 747–755.

Peron, R. M., & Allen, G. L. (1988). Attempts to train novices for beer flavor discrimination: A matter of taste. *The Journal of General Psychology,* 115, 403–418.

Piqueras-Fiszman, B. & Spence, C. (2016). Sensory expectations based on product-extrinsic food cues: An interdisciplinary review of the empirical evidence and theoretical accounts. *Food Quality and Preference*, 40, 165-179.

Ramírez, G., Hough, G., & Contarini, A. (2007). Influence of temperature and light exposure on sensory shelf-life of a commercial sunflower oil. Journal of Food Quality, 24, 195-204.

Risvik, E., McEwan, J. A., & Rodbotten, M. (1997). Evaluation of sensory profiling and projective mapping data. *Food Quality and Preference*, 8, 63–71.

Roberts, A.K. & Vickers, Z.M. (1994). A comparison of trained and untrained judges, evaluation of sensory attribute intensities and liking of Cheddar cheeses. *Journal of Sensory Studies*, 9, 1–20.

730 Rousseau, B. (2015). Sensory discrimination testing and consumer relevance. *Food Quality*
731 *and Preference,* 43, 122-125.
732 Saint-Eve, A., Lenfant, F., Teillet, E., Pineau, N., & Martin, N. (2011). Impact of panel training,
733 attribute list, type of response and dominance definition on TDS response. Poster
734 presentation at the *9th Pangborn Sensory Science Symposium*, Toronto, Canada.
735 Sawyer, F.M., Cardello, A.V. & Prell, P.A. (1988). Consumer evaluation of the sensory
736 properties of fish. *Journal of Sensory Studies,* 53, 12–18.
737 Schifferestein, H.N.J. (1996). Cognitive factors affecting taste intensity judgements. *Food*
738 *Quality and Preference,* 7, 167-175.
739 Sloman. S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*,
740 119, 30-52.
741 Stone, H. & Sidel, J. L. (2004). *Sensory Evaluation Practices* (Third Edition). San Diego:
742 Academic Press.
743 Stone, H. (2015) Alternative methods of sensory testing: advantages and disadvantages. In:
744 Rapid Sensory Profiling Techniques and Related Methods Applications in New Product
745 Development and Consumer Research. Edited by Delarue, J. Lawlor, JB. and Rogeaux,
746 M.. Woodhead Publishing, Cambridge. ISBN 978-1-78242-248-8
747 Teillet, E., Schlich, P., Urbano, C., Cordelle, S., & Guichard, E. (2010). Sensory methodologies
748 and the taste of water. *Food Quality and Preference*, 21, 967–976.
749 Torri, L., Dinnella, C., Recchia, A., Naes, T., Tuorila, H., & Monteleone, E. (2013). Projective
750 Mapping for interpreting wine aroma differences as perceived by naïve and experienced
751 assessors. Food Quality and Preference, 29(1), 6–15.
752 Valentin, D., Chollet, S., Lelièvre, M., & Abdi, H. (2012). Quick and dirty but still pretty good:
753 A review of new descriptive methods in food science. *International Journal of Food*
754 *Science and Technology*, 47, 1563–1578.
755 van't Riet, J., Sijtsema, SJ. Dagevos, H., de Bruijin, G.J. (2011) The importance of habits in
756 eating behaviour. An overview and recommendations for future research. Appetite, 57 pp.
757 585–596

758 Varela, P., & Ares, G. (2012). Sensory profiling, the blurred line between sensory and
759 consumer science. A review of novel methods for product characterization. *Food*
760 *Research International*, 48, 893–908.
761 Varela, P., Vidal, L., Antúnez, L., Silva-Cadena, R., Giménez, A. & Ares, G. (2014) What is
762 behind a napping test? Cognitive aspects of holistic approaches. Oral presentation, 6th
763 European Conference in Sensory and Consumer Research. Copenhagen, Denmark.
764 Veinand, B., Godefroy, C., Adam, C., & Delarue, J. (2011). Highlight of important product
765 characteristics for consumers. Comparison of three sensory descriptive methods
766 performed by consumers. *Food Quality and Preference*, 22, 474–485.
767 Veramendi, M., Herencia, P., & Ares, G. (2013). Perfume odor categorization: To what extent
768 trained assessors and consumer agree? *Journal of Sensory Studies*, 28, 76-89.
769 Vidal, L., Cadena, R.S., Ábalos, R.A., Gómez, B., Giménez, A., Varela, P., et al. (2013).
770 Assessment of global and individual reproducibility of projective mapping with consumers.
771 *Journal of Sensory Studies,* 29, 74-87.
772 Vidal, L., Silva Cadena, R., Antúnez, L., Giménez, A., Varela, P. & Ares, G. (2014). Stability
773 of sample configurations from projective mapping: How many consumers are necessary?
774 *Food Quality and Preference,* 34, 79–87
775 Vidal, L., Antúnez, L., Giménez, A., Varela, P., Deliza, R., Ares, G (2016) Can consumer
776 segmentation in projective mapping contribute to a better understanding of consumer
777 perception? Food Quality and Preference, 47, 64–72
778 Vidal, L., Jaeger, S.R., Antúnez, L., Giménez, A., & Ares, G. (2016). Product spaces derived
779 from projective mapping and CATA questions: Influence of replicated assessments and
780 increased number of study participants. *Journal of Sensory Studies*, In press.
781 Wolters, C.J. & Allchurch, E.M. (1994). Effect of training procedure on the performance of
782 descriptive panels. *Food Quality and Preference*, 5, 203–214.
783 Worch, T., Lê, S., & Punter, P. (2010). How reliable are the consumers? Comparison of
784 sensory profiles from consumers and experts. *Food Quality and Preference*, 21, 309–318.

785 Zamora, M. C. & Guirao, M. (2004), performance comparison between trained assessors and
786 wine experts using specific sensory attributes. *Journal of Sensory Studies,* 19: 530–545.